

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Satoshi Shinpuku et al.

Application. No.: to be assigned

Filed: March 12, 2004

Title: TEXT AND QUESTION GENERATING APPARATUS AND METHOD

:
:
:
:
:
:
:

Group Art Unit: to be assigned

Examiner: to be assigned

CLAIM FOR PRIORITY

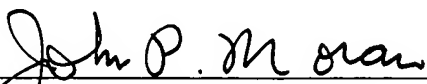
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Sir:

A certified copy of corresponding Japanese Application No. 2003-069658, filed March 14, 2003 is attached. It is requested that the right of priority provided by 35 U.S.C. 119 be extended by the U.S. Patent and Trademark Office.

Respectfully submitted,

Date: March 12, 2004



John P. Moran, Reg. No. 30,906
Swidler Berlin Shereff Friedman, LLP
3000 K Street, NW, Suite 300
Washington, DC 20007-5116
Telephone: (202) 424-7500
Facsimile: (202) 295-8478

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 3 月 1 4 日
Date of Application:

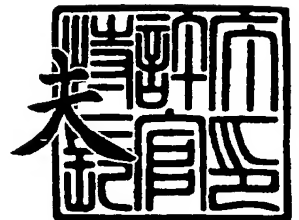
出 願 番 号 特 願 2 0 0 3 - 0 6 9 6 5 8
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 0 6 9 6 5 8]

出 願 人 富 士 通 株 式 会 社
Applicant(s):

2 0 0 3 年 1 2 月 2 2 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康



出証番号 出証特 2 0 0 3 - 3 1 0 6 2 7 2

【書類名】 特許願

【整理番号】 0350131

【提出日】 平成15年 3月14日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/30

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

 【氏名】 新福 哲

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

 【氏名】 難波 功

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100074066

 【弁理士】

 【氏名又は名称】 本間 崇

 【電話番号】 03-5224-3802

【手数料の表示】

 【予納台帳番号】 016713

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9704375



【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 テキスト情報作成装置、事例寄せ装置、FAQ作成用質問事例抽出装置、検索装置

【特許請求の範囲】

【請求項 1】 文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される属性入力部と、

前記節に付与され得る、談話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する属性である節長さ比率属性を作成する談話構造属性作成部と、

前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、を任意に組み合わせた属性である組合属性を作成する組合属性作成部と、

前記属性入力部に入力された人為属性と前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する重要度推定部と、

テキストを入力されるテキスト入力インタフェースと、

前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記重要度推定部で推定された各属性の重要度に基づいて決定する重要節決定部と、

前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力するテキスト出力インタフェースと、

を備えることを特徴とするテキスト情報作成装置。

【請求項 2】 文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される属性入力部と、

前記節に付与され得る、談話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する

属性である節長さ比率属性を作成する談話構造属性作成部と、

単語に関する属性である単語属性を作成する単語属性作成部と、

前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、前記単語属性作成部で作成された単語属性と、を任意に組み合わせた属性である組合属性を作成する組合属性作成部と、

前記属性入力部に入力された人為属性と前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と前記単語属性作成装置で作成された単語属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する重要度推定部と、

テキストを入力されるテキスト入力インタフェースと、

前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記重要度推定部で推定された各属性の重要度に基づいて決定する重要節決定部と、

前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力するテキスト出力インタフェースと、

を備えることを特徴とするテキスト情報作成装置。

【請求項 3】 文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される属性入力部と、

前記節に付与され得る、談話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する属性である節長さ比率属性を作成する談話構造属性作成部と、

前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、を任意に組み合わせた属性である組合属性を作成する組合属性作成部と、

前記属性入力部に入力された人為属性と前記談話構造属性作成部で作成された

談話構造属性および節長さ比率属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する重要度推定部と、

前記重要度推定部で重要度が推定された各属性の中から、余分と判断された属性である余分属性を削除する余分属性削除部と、

テキストを入力されるテキスト入力インタフェースと、

前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記余分属性削除部で削除されなかった属性の前記重要度推定部で推定された重要度に基づいて決定する重要節決定部と、

前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力するテキスト出力インタフェースと、

を備えることを特徴とするテキスト情報作成装置。

【請求項 4】 前記テキスト出力インタフェースが出力するテキストについての情報は、前記重要節決定部の決定で重要とされた節のみによって構成される要約文であることを特徴とする請求項 1～請求項 3 のいずれか 1 項に記載のテキスト情報作成装置。

【請求項 5】 請求項 1～請求項 4 のいずれか 1 項に記載のテキスト情報作成装置のテキスト出力インタフェースから出力される情報を利用して、テキスト群に存在する所望の内容が記載された複数のテキストを 1 つの集合にまとめることを特徴とする事例寄せ装置。

【請求項 6】 請求項 5 に記載の事例寄せ装置を使用して、複数の質問事例を少なくとも 1 つの質問事例集合に分類する手段と、

前記少なくとも 1 つの質問事例集合の中から、未来に質問されることが予測される質問事例を含む質問事例集合を決定する手段と、

前記決定した質問事例集合に含まれる質問事例を出力する手段と、

を有すること特徴とする F A Q (Frequently Asked Questions) 作成用質問事例抽出装置。

【請求項 7】 請求項 1～請求項 4 のいずれか 1 項に記載のテキスト情報作成装置のテキスト出力インタフェースから出力される情報を利用して、テキスト群の中から、所望の内容が記載されたテキストを検索することを特徴とする検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、テキストについての情報を作成するテキスト情報作成装置、および、このテキスト情報作成装置によって作成されるテキストについての情報を利用する事例寄せ装置、FAQ (Frequently Asked Questions) 作成用質問事例抽出装置、検索装置に関する。

【0002】

テキスト情報作成装置で作成されたテキストについての情報は、所望のテキストを複数のテキストの中から検索や事例寄せなどする際に利用される。

【0003】

ここで、検索とは、たとえば、指定された文や単語など（以下、「単語等」という。）と内容が類似している単語等を含むテキストを、複数のテキストの中から探し出すことをいう。

【0004】

また、事例寄せとは、たとえば、複数のテキストの中から、指定された要素や観点などが相互に類似しているテキストを探し出し、これらを 1 つのグループにまとめることをいう。

【0005】

【従来の技術】

テキストの検索や事例寄せなどを行う場合には、この検索や事例寄せなどを行うテキスト群における各テキストの内容を理解する必要があるが、その際、すべてのテキストのすべての箇所に目を通していたのでは、多大な時間や労力などを浪費してしまう。

【0006】

このため、従来は、この検索や事例寄せなどに伴う時間や労力などを軽減すべく、つぎの従来技術 1～従来技術 4 によってテキストについての情報を作成し、この作成された情報を利用して検索や事例寄せなどを行っていた。

【0007】

以下、この従来技術 1～従来技術 4 について簡単に説明する。

【0008】

(従来技術 1)

従来技術 1 は、検索や事例寄せなどを行うテキスト群とこのテキスト群を構成する各テキストとにおける単語の出現頻度を比較などすることにより、各テキストにおける単語を各テキストにおいて順位付けする技術である（特許文献 1 参照）。この従来技術 1 による順位を利用すれば、指定された単語が重要とされているテキストの検索や事例寄せなどが容易になる。

【0009】

(従来技術 2)

従来技術 2 は、各テキストについて談話構造解析を行い、各テキストにおける単語等に談話の種別たる談話構造を付与する技術である（特許文献 2 参照）。この従来技術 2 による談話構造を利用すれば、各テキストから、テキストの内容とはあまり関係がないと思われる単語等（たとえば、決まりきった挨拶文など）を除去することができるため、テキスト群における各テキストを調査する時間や労力が軽減し、検索や事例寄せなどが容易になる。

【0010】

(従来技術 3)

従来技術 3 は、テキスト群に含まれる各単語等に対して種別を付与し、この付与した種別ごとにデータベースを区分けする技術である（特許文献 2 参照）。この従来技術 3 により付与される種別を利用すれば、指定された単語等の種別と同じ種別の単語等を持つテキストと当該単語等を持たないテキストとを簡単に識別できるため、検索や事例寄せなどが容易になる。

【0011】

(従来技術 4)

従来技術 4 は、テキストの重要な単語を項目化するテンプレートを使用して、テキストを要約する技術である（特許文献 3 参照）。この従来技術 4 によれば、各テキストの要約を利用できるため、検索や事例寄せなどが容易になる。

【0012】

【特許文献 1】

特開平 8-305710 号公報

【特許文献 2】

特開 2002-278977 号公報

【特許文献 3】

特開 2002-24144 号公報

【0013】

【発明が解決しようとする課題】

しかし、検索や事例寄せなどにおいては、多くの場合、テキスト自体のベクトルが共通するテキストを探し出すことよりも、指定された単語等の内容を内容とするテキスト、または、指定された単語等の内容に類似する内容を内容とするテキストを探し出すことの方が重要である。

【0014】

したがって、上記従来技術 1～従来技術 4 には、それぞれ次のような問題があった。

以下、かかる問題を、図 23 に記載のテキスト群を用いて説明する。

【0015】

なお、図 23 に記載のテキスト群は、テキスト 1 とテキスト 2 とテキスト 3 とからなり、テキスト 1 は、文字面の点でテキスト 2 と類似し（なぜなら、「研修」や「鴨」といった文字が使われている点で、テキスト 1 とテキスト 2 とは共通するからである。）、内容面の点でテキスト 3 と類似する（なぜなら、「料理」に関する内容である点で、テキスト 1 とテキスト 3 とは共通するからである。）

。

【0016】

（従来技術 1 の問題）

従来技術 1 には、テキスト内の単語の順位を利用しても、指定された単語等と文字面で類似するテキストを容易に探し出すことできるだけで、内容面で類似するテキストを容易に探し出すことができないという問題があった。

【0017】

すなわち、たとえば図 23 においては、テキスト 1 とテキスト 2 とに「研修」や「鴨」といった珍しい単語が使用されているため、上記従来技術 1 で作成される順位を利用しても、テキスト 1 と内容面で類似するテキスト 3 をテキスト 1 に類似するテキストとして探し出すことが容易でないという問題があった。

【0018】

(従来技術 2 の問題)

また、従来技術 2 には、談話構造を利用しても、ある程度の余分な単語等しか除去できないため、残余の文字面による類似性がある程度重要視され、内容面で類似するテキストを必ずしも容易に探し出すことができないという問題があった。

【0019】

すなわち、たとえば図 23 においては、上記従来技術 2 に係る談話構造を利用しても、テキスト 1 と内容面で類似するテキスト 3 をテキスト 1 に類似するテキストとして探し出すことが容易でない可能性が高いという問題があった。

【0020】

(従来技術 3 の問題)

また、従来技術 3 には、単語に付与された種別情報を利用しても、指定された単語等と文字面において共通するが種別の異なる単語等を含んでいるテキストを容易に探し出すことができず、したがって、たとえば図 23 において、テキスト 1 と内容面において類似するテキスト 3 をテキスト 1 に類似するテキストとして探し出すことができないという問題があった。

【0021】

(従来技術 4 の問題)

そして、従来技術 4 では、談話などの多種多様な表現が混在するテキストの要約をテンプレートを使用して作成する場合、テキストから内容を抽出する際に使

用するテンプレートの型と各テンプレートを埋めるための条件とを作成するコストが膨大すぎるという問題があった。また、従来技術 4 には、テンプレートを事前に作成しきれないとまったくテンプレートを利用できないという問題があった。

【0022】

このように、従来技術 1～従来技術 4 によって作成される情報は、テキストの検索や事例寄せなどにおいて内容面で類似するテキストを探し出すための情報としては不十分であった。

【0023】

したがって、従来は、たとえば図 23 において、テキスト 1 と内容面において類似するテキスト 3 をテキスト 1 に類似するテキストとして探し出すことが極めて問題であった。

【0024】

そこで、本発明は、かかる事情に鑑み、テキストの内容に強く関連する単語等を過剰に膨大な量の人手によるコストを必要とせずにテキストから抽出し、この抽出した単語等を用いてテキストについての情報を作成するテキスト情報作成装置、および、このテキスト情報作成装置によって作成されるテキストについての情報を利用する事例寄せ装置、FAQ (Frequently Asked Questions) 作成用質問事例抽出装置、検索装置を提供することを目的とする。

【0025】

【課題を解決するための手段】

本発明によれば、上記課題は、次の手段によって解決される。

【0026】

第 1 の発明は、テキスト情報作成装置である。この第 1 の発明に係るテキスト情報作成装置は、属性入力部と、談話構造属性作成部と、組合属性作成部と、重要度推定部と、テキスト入力インタフェースと、重要節決定部と、テキスト出力インタフェースと、を備えることを特徴とする。ここで、属性入力部は、文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される。また、談話構造属性作成部は、前記節に付与され得る、談

話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する属性である節長さ比率属性を作成する。また、組合属性部は、前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、を任意に組み合わせた属性である組合属性を作成する。また、重要度推定部は、前記属性入力部に入力された人為属性と前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する。また、テキスト入力インタフェースは、テキストを入力される。また、重要節決定部は、前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記重要度推定部で推定された各属性の重要度に基づいて決定する。さらに、テキスト出力インタフェースは、前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力する。

【0027】

第 2 の発明は、テキスト情報作成装置である。この第 2 の発明に係るテキスト情報作成装置は、属性入力部と、談話構造属性作成部と、単語属性作成部と、組合属性作成部と、重要度推定部と、テキスト入力インタフェースと、重要節決定部と、テキスト出力インタフェースと、を備えることを特徴とする。ここで、属性入力部は、文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される。また、談話構造属性作成部は、前記節に付与され得る、談話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する属性である節長さ比率属性を作成する。また、単語属性作成部は、単語に関する属性である単語属性を作成する。また、組合属性作成部は、前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、前記単語属性作成部で作成された単語属性と、を任意に組み合わせた属性である組合属性を作成する。また、重要度推定部は、前記属性入力部に入力

された人為属性と前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と前記単語属性作成装置で作成された単語属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する。また、テキスト入力インタフェースは、テキストを入力される。また、重要節決定部は、前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記重要度推定部で推定された各属性の重要度に基づいて決定する。さらに、テキスト出力インタフェースは、前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力する。

【0028】

第 3 の発明は、テキスト情報作成装置である。この第 3 の発明に係るテキスト情報作成装置は、属性入力部と、談話構造属性作成部と、組合属性作成部と、重要度推定部と、余分属性削除部と、テキスト入力インタフェースと、重要節決定部と、テキスト出力インタフェースと、を備えることを特徴とする。ここで、属性入力部は、文書または文の一部である節に付与され得る、ユーザによって作成された属性である人為属性を入力される。また、談話構造属性作成部は、前記節に付与され得る、談話構造に関する属性である談話構造属性および前記節の文字数と前記節にマッチしたマッチングパターンの文字数との比率に関する属性である節長さ比率属性を作成する。また、組合属性作成部は、前記属性入力部に入力される人為属性と、前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と、を任意に組み合わせた属性である組合属性を作成する。また、重要度推定部は、前記属性入力部に入力された人為属性と前記談話構造属性作成部で作成された談話構造属性および節長さ比率属性と前記組合属性作成装置で作成された組合属性とについて、これら属性が前記節に付与された場合に前記節とテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する。また、余分属性削除部は、前記重要度推定部で重要度が推定された各属性の中から、余分と判断された属性である余分属性を削除する。また、テキスト入力インタフ

ェースは、テキストを入力される。また、重要節決定部は、前記テキスト入力インタフェースに入力されたテキストにおける 1 以上の節の中から、前記テキスト入力インタフェースに入力されたテキストの内容との相関が高い重要な節を、前記余分属性削除部で削除されなかった属性の前記重要度推定部で推定された重要度に基づいて決定する。さらに、テキスト出力インタフェースは、前記重要節決定部の決定に基づいて作成される前記テキスト入力インタフェースに入力されたテキストについての情報を出力する。

【0029】

第 4 の発明は、上記第 1 の発明～上記第 3 の発明のいずれか 1 つに係るテキスト情報作成装置である。この第 4 の発明に係るテキスト情報作成装置は、前記テキスト出力インタフェースが出力するテキストについての情報は、前記重要節決定部の決定で重要とされた節のみによって構成される要約文であることを特徴とする。

【0030】

第 5 の発明は、事例寄せ装置である。この第 5 の発明に係る事例寄せ装置は、上記第 1 の発明～上記第 4 の発明のいずれか 1 つに係るテキスト情報作成装置のテキスト出力インタフェースから出力される情報を利用して、テキスト群に存在する所望の内容が記載された複数のテキストを 1 つの集合にまとめることを特徴とする。

【0031】

第 6 の発明は、FAQ (Frequently Asked Questions) 作成用質問事例抽出装置である。この第 6 の発明に係る FAQ (Frequently Asked Questions) 作成用質問事例抽出装置は、上記第 5 の発明に係る事例寄せ装置を使用して、複数の質問事例を少なくとも 1 つの質問事例集合に分類する手段と、前記少なくとも 1 つの質問事例集合の中から、未来に質問されることが予測される質問事例を含む質問事例集合を決定する手段と、前記決定した質問事例集合に含まれる質問事例を出力する手段と、を有すること特徴とする。

【0032】

第 7 の発明は、検索装置である。この第 7 の発明に係る検索装置は、上記第 1

の発明～上記第4の発明のいずれか1つに係るテキスト情報作成装置のテキスト出力インタフェースから出力される情報を利用して、テキスト群の中から、所望の内容が記載されたテキストを検索することを特徴とする。

【0033】

【発明の実施の形態】

以下に、添付した図面を参照しつつ、本発明に係るテキスト情報作成装置の好適な実施の形態を詳細に説明する。

【0034】

図1は、本発明の実施の形態に係るテキスト情報作成装置の概要図である。

【0035】

図1に示すように、本発明の実施の形態に係るテキスト情報作成装置は、属性入力部と、単語属性作成部と、余分属性削除部と、組合属性作成部と、談話構造属性作成部と、重要度推定部と、重要節決定部と、テキスト入力インタフェースと、テキスト出力インタフェースと、を有している。

【0036】

また、本発明の実施の形態に係るテキスト情報作成装置には、属性集合用DBと、コーパスDBと、談話構造解析ルールDBと、結果DBと、重要度DBと、が接続されている。なお、DBとは、Data Base（データベース）の略である。また、コーパスとは言語資料体を意味し、コーパスDBには、テキストが大規模または網羅的に格納されている。

【0037】

本発明の実施の形態に係るテキスト情報作成装置は、テキスト入力インタフェースから入力されたテキストについての情報を作成し、この作成した情報をテキスト出力インタフェースから出力する。

【0038】

なお、テキストについて情報とは、たとえば、テキストの重要箇所を強調表示した情報やテキストの要約文などをいう。

【0039】

図2は、本発明の実施の形態に係るテキスト情報作成装置における処理を説明

するフローチャートである。

【0040】

まず、本発明の実施の形態に係るテキスト情報作成装置においては、前処理が行われる（ステップS2-1）。

【0041】

ここで、前処理とは、節（テキストに記載されている文章の一部またはこの文章を構成する文の一部をいう。テキストに記載されている文章またはこの文章を構成する文は、少なくとも1つの節によって構成される。）に付与が可能な少なくとも1つの属性を作成しまたは入力される処理と、この作成しまたは入力された属性の重要度を推定する処理と、上記作成しまたは入力された属性と上記作成しまたは入力された属性の重要度との対応関係を図14の属性重要度用DBに書き込む処理と、からなる。

【0042】

なお、上記したところからも明らかなように、属性集合とは、少なくとも1つの属性から構成される属性の集合である。

また、属性とは、テキスト情報作成装置によって節に付与される性質や特徴をいう。

【0043】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図1のテキスト入力インタフェースから入力されたテキストが読み込まれる（ステップS2-2）。

【0044】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS2-2で読み込んだテキストを構成する各節の重要度がそれぞれ推定され、この推定された各節の重要度に基づいて各節が重要である否かが決定され、節および節の重要度および節の重要・非重要な別が図13の結果DBに書き込まれる（ステップS2-3）。

【0045】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図13

の結果DBに書き込まれた決定が重要となっている節（重要節）のみをテキスト出力インタフェースから出力する否かが判断される（ステップS2-4）。

【0046】

ステップS2-4で、決定が重要となっている節（重要節）のみを出力すると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、決定が重要となっている節（重要節）のみを示す要約文が出力インタフェースから出力される（ステップS2-5）。たとえば、上記ステップS2-2で図15に記載のテキストが読み込まれた場合、本発明の実施の形態に係るテキスト情報作成装置からは、図17に記載のテキストが出力される。

【0047】

他方、ステップS2-4で、決定が重要となっている節（重要節）のみを出力しないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、決定が重要となっている節（重要節）のみを強調表示したテキストが出力インタフェースから出力される（ステップS2-6）。たとえば、上記ステップS2-2で図15に記載のテキストが読み込まれた場合、本発明の実施の形態に係るテキスト情報作成装置からは、図18に記載のテキストが出力される。

【0048】

図3は、図2のステップS2-1で行われる前処理を説明するフローチャートである。

【0049】

前処理においては、まず、本発明の実施の形態に係るテキスト情報作成装置において、少なくとも1つの属性が初期属性集合を構成する属性として作成される（ステップS3-1）。

【0050】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、初期属性集合に単語属性を加えるか否かが判断される（ステップS3-2）。

【0051】

ステップS3-2で、初期属性集合に単語属性を加えないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合

用DBの仮属性集合と余分含属性集合とが初期属性集合で上書される（ステップS3-3）。

【0052】

他方、ステップS3-2で、単語属性を加えると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、単語属性作成部によって単語属性の作成に関する処理が行われる（ステップS3-4）。

【0053】

上記ステップ3-4で単語属性作成部による単語属性の作成に関する処理が行われた場合、本発明の実施の形態に係るテキスト情報作成装置においては、このステップS3-4で図11の属性集合用DBの仮属性集合に単語属性が加えられたか否かが判断される（ステップS3-13）。

【0054】

上記ステップS3-13で仮属性集合に単語属性が加えられたと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBの仮属性集合を構成する単語属性の数が閾値以上であるか否かが判断される（ステップS3-14）。

【0055】

上記ステップS3-14で単語属性の数が閾値以上でないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS3-4に戻って単語属性作成部による単語属性の作成に関する処理が行われる。

【0056】

他方、上記ステップS3-14で単語属性の数が閾値以上であると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、下記ステップS3-5の処理が行われる。

【0057】

上記ステップS3-13で仮属性集合に単語属性が加えられなかったと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、下記ステップS3-5の処理が行われる。

【0058】

つぎに、本発明の実施の形態に係るテキスト情報作成装置は、余分属性を削除するか否かを判断する（ステップS3-5）。

【0059】

上記ステップS3-5で、余分属性の削除を行わないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBに格納されている余分除属性集合と仮属性集合とが余分含属性集合で上書される（ステップS3-6）。

【0060】

上記ステップS3-6で上書きがなされた場合、本発明の実施の形態に係るテキスト情報作成装置においては、最終確認が行われる（ステップS3-7）。

【0061】

他方、ステップS3-5で、余分属性の削除を行うと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、余分属性削除部によって余分属性が削除される（ステップS3-8）。

【0062】

上記ステップS3-8で余分属性が削除された場合、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS3-8で余分除属性集合が上書されたか否かが判断される（ステップS3-9）。

【0063】

上記ステップS3-9で、余分除属性集合が上書されなかったと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、ステップS3-5に戻って処理がやり直される。

【0064】

他方、上記ステップS3-9で、余分除属性集合が上書されたと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、最終確認が行われる（ステップS3-7）。

【0065】

上記ステップS3-7の最終確認が終了すると、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS3-7の最終確認で図11の属

性集合用DBの仮属性集合が上書されたか否かの判断が行われる（ステップS3-10）。

【0066】

上記ステップS3-10で、仮属性集合が上書されたと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、上記単語属性をさらに新たに加えるか否かの判断が行われる（ステップS3-11）。

【0067】

上記ステップS3-11で、単語属性を新たに加えるとの判断がなされた場合、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS3-4に戻って単語属性作成部による単語属性の作成に関する処理が行われる。

【0068】

他方、上記ステップS3-11で、単語属性を新たに加えないとの判断がなされた場合、本発明の実施の形態に係るテキスト情報作成装置においては、ステップS3-5に戻って余分属性を削除するか否かが判断される。

【0069】

上記ステップS3-10で、仮属性集合が上書されなかったと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図14の重要度推定部によって、図11の属性集合用DBの最終属性集合を構成する各属性の重要度の推定が行われ、この推定された重要度が図14の重要度DBに書き込まれる（ステップS3-12）。

【0070】

上記ステップ3-12で各属性の重要度が推定され、この推定された重要度が図14の重要度DBに書き込まれた場合、本発明の実施の形態に係るテキスト情報作成装置においては、図2のステップS2-1の前処理が終了する。

【0071】

図4は、図3のステップS3-1で行われる初期属性集合を構成する属性の作成を説明するフローチャートである。

【0072】

初期属性集合を構成する属性を作成する場合、本発明の実施の形態に係るテキ

スト情報作成装置においては、まず、図12のコーパスDBから正解内容と節とが読み込まれる（ステップS4-1）。

【0073】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図12のコーパスDBから読込んだ各節に対して談話構造属性作成部によって談話構造解析が行われる（ステップS4-2）。

【0074】

この談話構造解析においては、まず、図10の談話構造解析ルールDBのマッチングパターンとテキストを構成する各節とのマッチングが行われる。なお、図10の談話構造解析ルールDBのマッチングパターンは、あらかじめ作成されている。

【0075】

図10の談話構造解析ルールDBのマッチングパターンと節とがマッチしたら、このマッチした節は、マッチしたマッチングパターンに対応する談話構造だと決定され、この決定された談話構造とマッチ文字数（マッチングパターンの文字数）とが示された談話タグが節ごとに付与される。

【0076】

この談話構造解析によると、たとえば、談話構造属性作成部において、図15で示すテキストが入力された場合、図16で示すテキストが出力される。

【0077】

なお、1つの節の同じ部分に2つのマッチングパターンがマッチした場合は、図10の談話構造ルールDBにおいて上位に記載されているマッチングパターンが優先されるため、1つの節の同じ部分にいくつもの談話構造が付与されることはない。たとえば、ある1つの節に、図10の談話構造ルールDBのマッチパターン「していただけますか」と「くださいますか」とがマッチした場合は、図10の談話構造ルールDBにおいて上位に記載されている「していただけますか」が優先される。したがって、この場合は、この「していただけますか」というマッチパターンが、上記ある1つの節にマッチするとされる。ただし、「～ですが、～できません」であれば、「～ですが」にマッチした談話構造と「～できませ

ん」にマッチした談話構造とを1つの節の異なる部分にもたせることが可能である。

【0078】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図12のコーパスDBの各節についてそれぞれ付与された各談話構造が、図11の属性集合用DB内の初期属性集合に談話構造属性として書き込まれる（ステップS4-3）。

【0079】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図12のコーパスDBの各節についてそれぞれ付与された各マッチ文字数と節の文字数との比率が、図12のコーパスDBに格納されている各節について、それぞれ算出される（ステップS4-4）。

【0080】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、算出された各比率についてクラスタリング処理（たとえば、整数単位が同じであるというような近い数値にある比率を一つの比率で表す処理など）を施すか否かを判断する（ステップS4-5）。

【0081】

本発明の実施の形態に係るテキスト情報作成装置においては、このステップS4-5におけるクラスタリング処理を行うか否かの判断を、たとえば、データスパースネスの問題（後述する機械学習時に使用可能なデータが希薄すぎるという問題）が生じないか否かに基づいて行うことができる。

【0082】

上記ステップS4-5で、各比率にクラスタリング処理を行うと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、各比率についてクラスタリング処理が行われ、クラスタリング処理後の各比率が、節長さ比率属性として、図11の属性集合用DB内の初期属性集合に書き込まれる（ステップS4-6）。

【0083】

他方、上記ステップS4-5で、各比率にクラスタリング処理を行わないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、各節について算出された、マッチ文字数と節の文字数との比率が、節長さ比率属性として、図11の属性集合用DB内の初期属性集合に書き込まれる（S4-7）。

【0084】

上記ステップS4-6または上記ステップS4-7で、節長さ比率属性が図11の属性集合用DB内の初期属性集合に書き込まれた場合、本発明の実施の形態に係るテキスト情報作成装置においては、ユーザによって作成された属性が属性入力部を介して読み込まれる（ステップS4-8）。なお、ユーザは、所望の単語や文などを属性として任意に作成し、入力することができる。

【0085】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、属性入力部を介して読み込まれた属性のうち、図12のコーパスDBに現れない属性を削除する（ステップS4-9）。

【0086】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、読み込まれた属性のうち、上記ステップS4-9で削除されなかった属性が、人為属性として、図11の属性集合用DB内の初期属性集合に書き込まれる（ステップS4-10）。

【0087】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、属性集合用DBの初期属性集合が読み込まれる（ステップS4-11）。

【0088】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、組合属性作成部によって、図11の属性集合用DBの初期属性集合における各属性を2つ以上組み合わせたものが、組合属性として作成される（ステップS4-12）。

【0089】

たとえば、「比率が2倍以上である」という節長さ比率属性と「解決という文字がある」という人為属性とを組み合わせると、「比率が2倍以上であり且つ解決という文字がある」という組合属性が作成される。また、たとえば、「談話構造が質問である」という談話構造属性と「比率が2倍以下である」という節長さ比率属性とを組み合わせると、「談話構造が質問であり且つ比率が2倍以下である」という組合属性が作成される。

【0090】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS4-12で作成された組合属性が、図11の属性集合用DBの初期属性集合に書き加えられる（ステップS4-13）。

【0091】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBにおいて、仮属性集合と確認属性集合とが初期属性集合で上書される（ステップS4-14）。

【0092】

図5は、図3のステップS3-4で行われる、単語属性作成部による単語属性の作成に関する処理を説明するフローチャートである。

【0093】

単語属性作成部による単語属性の作成に関する処理を行う場合、本発明の実施の形態に係るテキスト情報作成装置においては、まず、単語属性作成部によって、図12のコーパスDBから節と正解内容とが読み込まれる（ステップS5-1）。

【0094】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、単語属性作成部によって、図11の属性集合用DB内の仮属性集合が読み込まれる（ステップS5-2）。

【0095】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、単語属性作成部によって、読み込まれた仮属性集合で図11の属性集合用DB内の最終

属性集合が上書される（ステップS5-3）。

【0096】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、重要度推定部により、図11の属性集合用DBの最終属性集合を構成する各属性の重要度が推定される（ステップS5-4）。

【0097】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、重要節決定部によって、図12のコーパスDBに格納される各節について重要度が決定され、この決定の結果が図13の結果DBに書き込まれる（S5-5）。

【0098】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図13の結果DBに書き込まれた決定で、図12のコーパスDBの試験決定が上書される（ステップS5-6）。

【0099】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、ステップS5-6で図12のコーパスDBのすべての試験決定が上書されたか否かを判断する（ステップS5-7）。

【0100】

上記ステップS5-7で図12のコーパスDBのすべての試験決定が上書されていないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、上記ステップS5-5に戻る。

【0101】

他方、上記ステップS5-7で図12のコーパスDBのすべての試験決定が上書されたと判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、図12のコーパスDBから、正解決定と試験決定とが異なる節をすべて読み込む（ステップ5-8）。

【0102】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、読み込まれたすべての節に閾値以上の頻度で出現する単語があるか否かを判断する（ス

テップ S 5-9)。

【0103】

上記ステップ S 5-9 で閾値以上の頻度で出現する単語がないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図 11 の属性集合用 DB の仮属性集合で余分含属性集合が上書される (ステップ S 5-15)。

【0104】

他方、上記ステップ S 5-9 で閾値以上の頻度で出現する単語があると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、この閾値以上の頻度で出現する単語のうち、頻度の一番高い単語が抽出される (ステップ S 5-10)。

【0105】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、ステップ S 5-10 で抽出した単語が、図 11 の属性集合用 DB の仮属性集合に既に存在しているか否かが判断される (ステップ S 5-11)。

【0106】

上記ステップ S 5-11 において、上記ステップ S 5-10 で抽出した単語が図 11 の属性集合用 DB の仮属性集合に既に存在していると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、図 11 の属性集合用 DB の仮属性集合で余分含属性集合が上書される (ステップ S 5-15)。

【0107】

他方、上記ステップ S 5-11 において上記ステップ S 5-10 で抽出した単語が図 11 の属性集合用 DB の仮属性集合に既に存在していないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、上記抽出した単語が図 11 の属性集合用 DB における初期属性集合に書き加えられる (S 5-12)。

【0108】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、組合属性作成部によって、図 11 の属性集合用 DB の初期属性集合を構成する各属性が

組み合わせられ、組合属性が作成される（ステップS5-13）。

【0109】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBの初期属性集合を構成する各属性のうち仮属性集合にない属性と上記ステップS5-13で作成した組合属性のうち仮属性集合にない属性とが図11の属性集合用DBの仮属性集合に書き加えられる（ステップS5-14）。

【0110】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBの仮属性集合で余分含属性集合が上書される（ステップS5-15）。

【0111】

図19は、単語属性を加えていない場合の属性集合である初期属性集合を用いた場合の、コーパス中のテキストIDが2のテキストを重要度決定器に入力した際の各節と各重要度とを示す図であり、図20は、単語属性を加えた場合の属性集合を用いた場合の、コーパス中のテキストIDが2のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。

【0112】

図19、図20をみれば明らかなように、PC、突然、設定などの単語に関する属性が加えられたことで、重要度が2番目に高い節も変更され、精度が増しているのがわかる。

【0113】

図6は、図3のステップ3-8で行われる余分属性削除部による処理を説明するフローチャートである。

【0114】

余分属性削除部による処理を行う場合、本発明の実施の形態に係るテキスト情報作成装置においては、まず、図11の属性集合用DBの仮属性集合が読み込まれる（ステップS6-1）。

【0115】

つぎに、本発明の実施の形態に係るテキスト情報作成装置は、図 11 の属性集合用 DB において、仮属性集合を構成する各属性を最終属性集合に書き加える（ステップ S 6-2）。

【0116】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、重要度推定部によって、図 11 の属性集合用 DB の最終属性集合に含まれる各属性の重要度が推定される（ステップ S 6-3）。

【0117】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図 11 の属性集合用 DB の最終属性集合に、重要度が閾値以下の属性が無いかが判断される（ステップ S 6-4）。

【0118】

つぎに、上記ステップ S 6-4 で重要度が閾値以下の属性があると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、重要節決定部によって、図 12 のコーパス DB の各テキストを構成する各節の重要度が決定され、出力が図 13 の結果 DB に書き込まれる（ステップ S 6-5）。

【0119】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図 13 の結果 DB に基づいて図 12 のコーパス DB の試験決定が上書される（ステップ S 6-6）。

【0120】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップ S 6-6 で、図 12 のコーパス DB の全テキストの試験決定が上書されたか否かを判断する（ステップ S 6-7）。

【0121】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図 14 の重要度 DB から各属性と各属性の重要度とを読み込んで、重要度が最も低い属性を選び出し、選ばれた重要度が最も低い属性を図 11 の属性集合用 DB の余分含属性集合から削除する（ステップ S 6-8）。

【0122】

ここで、上記ステップS6-8で選ばれる属性は、選ばれた属性が節に入っていた場合に節が重要でないというマイナスの重要度を持つ属性ではなく、選ばれた属性が節に入っている場合に節が重要であるまたは重要でないことを示す重要度を持つ属性である。

【0123】

上記ステップS6-8で選ばれる属性は、たとえば、最大エントロピー法による学習法を例にとると、属性が節に含まれる場合に重要な節であるという重みと、属性が節に含まれる場合に重要な節でないという重みが半々の割合となる属性である。

【0124】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性集合用DBにおいて、余分含属性集合が最終属性集合に書き込まれる（ステップS6-9）。

【0125】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、重要度推定部によって、図11の属性集合用DBの最終属性集合を構成する各属性の重要度が推定される（ステップS6-10）。

【0126】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、重要節決定部によって、図12のコーパスDBの各節の重要度が決定され、この決定の結果が図13の結果DBに書き込まれる（ステップS6-11）。

【0127】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図13の結果DBに基づいて図12のコーパスDBの余分除決定が上書される（ステップS6-12）。

【0128】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、上記ステップS6-13で図12のコーパスDBのすべての余分除決定が上書されたか

否かが判断される（ステップ S 6-13）。

【0129】

上記ステップ S 6-13 において図 12 のコーパス DB のすべての余分除決定が上書されたと判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、図 12 のコーパス DB の正解決定と試験決定と余分除決定とを比較して正解率を算出する（ステップ S 6-14）。

【0130】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、試験決定と正解決定とを、余分除決定と正解決定とを各々比較し、試験決定と正解決定とが一致する回数を求め、この一致する回数に決められた閾値を足した数値を算出し、この算出した数値よりも、余分除決定と正解決定とが一致する回数が多いか否かを判断する（ステップ S 6-15）。

【0131】

上記ステップ S 6-15 で多いと判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、図 11 の属性集合用 DB において、最終属性集合で仮属性集合と余分含属性集合とを上書きする（ステップ S 6-16）。

【0132】

他方、上記ステップ S 6-15 で少ないと判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、図 11 の属性集合用 DB において、余分属性を除く前の余分含属性集合で余分除属性集合を上書きする（ステップ S 6-17）。

【0133】

なお、図 21 は、余分属性を削除していない場合の属性集合である余分含属性集合を用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図であり、図 22 は、余分属性を削除した場合の属性集合である余分除属性集合を用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。図 21、図 22 を見ればわかるように、属性を削除したにもかかわらず、精度が同程度であることがわかる。

【0134】

このように、余分属性の削除を行うと、精度を同程度に保ちつつ属性の量が減少しているため、図2の下部にある、実際の入力きた際の実行速度を向上させることができるというメリットがある。

【0135】

図7は、図3のステップS3-7で行われる最終確認を説明するフローチャートである。

【0136】

最終確認を行う場合、本発明の実施の形態に係るテキスト情報作成装置においては、まず、図11の属性選択用DBから余分除属性集合が読み込まれる（ステップS7-1）。

【0137】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図11の属性選択用DBから確認属性集合が読み込まれる（ステップS7-2）。

【0138】

つぎに、本発明の実施の形態に係るテキスト情報作成装置は、図11の属性集合用DBにおいて、確認属性集合と余分除属性集合とが同じ属性集合であるか否かを判断する（ステップS7-3）。

【0139】

上記ステップS7-3で異なる属性集合であると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、この異なる属性であるという判断が閾値回数以上行われたか否かが判断される（ステップS7-4）。

【0140】

上記ステップS7-3で確認属性集合と余分除属性集合とが同じであると判断された場合または上記ステップS7-4で閾値以上行っていると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、余分除属性集合で最終属性集合が上書される（ステップS7-8）。

【0141】

上記ステップS7-4で閾値未満であると判断された場合、本発明の実施の形

態に係るテキスト情報作成装置においては、余分除属性集合で仮属性集合が上書きされる（ステップS7-6）。

【0142】

図8は、図5のステップS5-4で行われる重要度推定部の処理を説明するフローチャートである。

【0143】

重要度推定部の処理においては、本発明の実施の形態に係るテキスト情報作成装置は、まず、図11の属性集合用DBの最終属性集合を読み込む（ステップS8-1）。

【0144】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図12のコーパスDBから各節と各正解内容とが読み込まれる（ステップS8-2）。

【0145】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、読み込まれた各節と各正解内容の重要度とに基づいて機械学習が行われ、図11の属性集合用DBの最終属性集合に含まれる各属性の重要度が推定される（ステップS8-3）。

【0146】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図14の重要度用DBのデータがすべて消され、図11の属性集合用DBの最終属性集合の各属性と上記ステップS8-3で推定した各属性の重要度とが図14の重要度用DBに書き込まれる（ステップS8-4）。

【0147】

なお、上記ステップS8-3における機械学習の手法としては、各属性の重要度となる数値または度合を示す表現が推定できるならば、どのような機械学習の手法でも用いることができる。

【0148】

たとえば、最大エントロピー法（「言語と計算-4 確率的言語モデル」、東京大学出版会、P158）および最大エントロピー法の内部パラメータ推定法で

ある反復スケーリング法（「言語と計算－4 確率的言語モデル」、東京大学出版会、P 163）を用いて、各属性を属性が含まれている節が重要であるか重要でないかを示す素性関数 $\{F(\text{重要}|\text{属性}), F(\text{非重要}|\text{属性})\}$ の組と考え、各属性に関する各素性関数 $F(\cdot)$ の重みを上記反復スケーリング法を用いて推定することで、属性毎の重要度を推定する手法がある。

属性毎重要度式の一例を数 1 に示す。

【0149】

【数 1】

$$\frac{\exp^{\lambda_i F(\text{重要}|\text{属性})}}{\exp^{\lambda_i F(\text{重要}|\text{属性})} + \exp^{\lambda_j F(\text{非重要}|\text{属性})}}$$

他にも、各属性が単純にコーパスの内容が重要となっている節内と重要になっていない節内とに出てきた回数、およびベイズの定理（「言語と計算－4 確率的言語モデル」、東京大学出版会、P 4）から、属性が節に含まれている時に節が重要である条件付確率 $P(\text{重要}|\text{属性})$ を算出する手法などが考えられる。

【0150】

図 9 は、重要節決定部による処理を説明するフローチャートである。

【0151】

重要節決定部による処理においては、まず、本発明の実施の形態に係るテキスト情報作成装置において、図 11 の属性集合用 DB から最終属性集合が読み込まれる（ステップ S 9-1）。

【0152】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図 1 のテキスト入力インタフェースからテキストが読み込まれる（ステップ S 9-2）。

。

【0153】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、入力さ

れたテキストの文章または文を構成する各節に、図 11 の属性集合用 DB の最終属性集合に含まれる属性が付与される（ステップ S 9-3）。

【0154】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、図 14 の重要度 DB から付与された各属性の重要度が読み込まれる（ステップ S 9-4）。

【0155】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、上記読み込まれた各属性の重要度に基づいて、入力されたテキストの文章または文を構成する各節の重要度が推定される（ステップ S 9-5）。

【0156】

各節の重要度を推定する方法は、上記重要度推定部で行われる機械学習の手法と推定された属性毎の重要度の形態とにより様々であるが、たとえば、上記の最大エントロピー法によって各属性に関する二つの素性関数の重みを推定していた場合の推定方法の例として、次のような方法を採用することができる。

【0157】

すなわち、各節に対して、属性集合を構成する各属性の素性関数の組から、各属性が属性集合に存在する場合に節が重要であることを示す素性関数の重みを掛け合わせることで算出される数値と、各属性が属性群に存在する場合に節が重要でないことを示す素性関数の重みを各々掛け合わせることで算出される数値と、の割合を重要度とする方法がある。

【0158】

つぎに、本発明の実施の形態に係るテキスト情報作成装置においては、テキストが複数の節から構成されているかを判断する（ステップ S 9-6）。

【0159】

つぎに、上記ステップ S 9-6 で単数の節からなっていると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、決定を重要として、決定、節および算出した重要度を図 14 の結果 DB に書き込む（ステップ S 9-12）。

【0160】

他方、上記ステップS9-6で複数の節からなっていると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、変数Nが2に設定される（ステップS9-7）。

【0161】

つぎに、変数Nが2以上の数である場合、本発明の実施の形態に係るテキスト情報作成装置においては、重要度がN番目に高い節について、重要度があらかじめ定められた閾値以上であるか否かが判断される（ステップS9-8）。

【0162】

上記ステップS9-8で重要度がN番目に高い節の重要度が閾値以上であると判断された場合、本発明の実施の形態に係るテキスト情報作成装置においては、変数Nが、テキストを構成する節数ごとにあらかじめ定められている閾値以下であるか否かが判断される（ステップS9-9）。

【0163】

上記ステップS9-9で変数Nがテキストを構成する節数ごとにあらかじめ定められている閾値以下であると判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、変数Nを1増やし（ステップS9-10）、ステップS9-8に戻る。

【0164】

他方、上記ステップS9-9で変数Nがテキストを構成する節数ごとにあらかじめ定められている閾値以上であると判断された場合、本発明の実施の形態に係るテキスト情報作成装置は、すべての決定を非重要とした後、テキストに含まれる節数毎に決められた閾値数までの決定を重要度が高い順に重要と変更し、テキスト内の全決定、全節及び全内容を結果DBに書き込む（ステップS9-11）。

。

【0165】

すなわち、ステップS9-11では、重要度が高い順にN-1個の節の決定が重要とされ、他が非重要とされ、テキスト内の全決定、全節及び全内容が結果DBに書き込まれる。

【0166】

以上説明した本発明の実施の形態に係るテキスト情報作成装置によれば、たとえば、図15に記載のテキストを入力とした時に、図15のテキストにおける重要節のみで構成される図17に記載の要約文（たとえば、メールの文章を1～3節程度に要約した文）を出力することができるほか、図15のテキストにおいて重要節のみを強調表示した図18に記載のテキストを出力することができる。

【0167】

したがって、本発明の実施の形態に係るテキスト情報作成装置が出力する情報を用いれば、検索や事例寄せなどのテキストの類似性の検討を必要とする作業や処理などを、容易に行うことができる。

【0168】

この本発明の実施の形態に係るテキスト情報作成装置は、たとえば、次の実施例のように、用いることもできる。

【0169】

<実施例1>

本発明の実施の形態に係るテキスト情報作成装置の実施例1は、本発明の実施の形態に係るテキスト情報作成装置を含み、本発明の実施の形態に係るテキスト情報作成装置が出力する要約文に基づいて、所望の内容を有する複数の事例を1つの集合にまとめる事例寄せ装置である。

【0170】

実施例1に係る事例寄せ装置は、複数の事例がそれぞれ記載された複数のテキストがある場合に、これらを本発明の実施の形態に係るテキスト情報作成装置に入力して、その出力の類似するテキストを1つの集合にまとめる。

【0171】

出力が類似するか否かを判断する手法は、特に規定しないが、たとえばベクトル空間法（参考論文：Addison-Wesley Publishing（1989），Automatic Text Processing, pp. 312-325, Salton, G. : The Vector Space Modelを参照）で使用されている手法を用いればよい。

【0172】

以下、実施例 1 に係る事例寄せ装置を、図 23 のテキスト 1、テキスト 2 およびテキスト 3 を使用して、具体的に説明する。

【0173】

実施例 1 に係る事例寄せ装置は、テキスト 1、テキスト 2、テキスト 3 を入力とした時の本発明の実施の形態に係るテキスト情報作成装置の各出力に基づいて、テキスト内の単語に関して方向ベクトルを作成し、各ベクトル間の距離をベクタースペースモデルの手法を用いて算出する（ここでは説明のため、距離が 1 の場合を距離が最近とし、距離が 0 の場合を距離が最長として話を進める）。

【0174】

ここで、実施例 1 に係る事例寄せ装置において、テキスト 1 の要約文とテキスト 2 の要約文とのベクトル間の距離の絶対値が 0.8、テキスト 1 の要約文とテキスト 3 の要約文のベクトル間の距離の絶対値が 0.95、テキスト 2 の要約文とテキスト 3 の要約文のベクトル間の距離の絶対値が 0.82 と算出されたとすると、実施例 1 に係る事例寄せ装置は、テキスト 1 がテキスト 2 よりもテキスト 3 に近いと判断し、たとえば、まとめる閾値が 0.88 ならばテキスト 1 とテキスト 3 を同じ系統のテキストとしてまとめることができ、テキスト 1 とテキスト 2 とを、また、テキスト 2 とテキスト 3 とを、1 つの組にまとめないということができる。

【0175】

<実施例 2>

本発明の実施の形態に係るテキスト情報作成装置の実施例 2 は、実施例 1 に係る事例寄せ装置を含んだ、FAQ (Frequently Asked Questions) 作成用質問事例抽出装置である。

【0176】

実施例 2 に係る FAQ 作成用質問事例抽出装置は、実施例 1 に係る事例寄せ装置を使用して、複数の質問事例が格納されている DB に対して事例寄せを行い、複数の質問事例をいくつかの質問事例集合に分類する。

【0177】

そして、実施例 2 に係る F A Q 作成用質問事例抽出装置は、各質問事例集合のうち、未来に質問されることが予測される質問事例を含む質問事例集合を決定し、この決定した質問事例集合に含まれる質問事例を出力する。

【0178】

未来に来ることが予測される質問事例を含む質問事例集合を決定する手法については、特に言及しないが、たとえばテキスト数が多い質問事例集合や最近頻繁に質問が寄せられた質問事例を含む質問事例集合を選ぶ方法が考えられる。

【0179】

出力となる質問事例集合の質問事例を決定する手法についても特に言及しないが、たとえば上記事例寄せ装置がベクタスペースモデルの手法を用いた際、集合内で中心的な位置を示すベクトルを持つテキスト自身または中心的な位置を示すベクトルを持つテキストを本発明の実施の形態に係るテキスト情報作成装置の入力とし、この場合の出力を使用する手法が考えられる。

【0180】

たとえば、図 23 に記載されている 3 つのテキストと同様のテキストが大量に DB に存在し、テキスト 1 の要約文のベクトルが中心を示すテキストの集合が存在した場合、テキスト 1 の内容が F A Q 作成用の質問事例として出力される。

【0181】

<実施例 3>

本発明の実施の形態に係るテキスト情報作成装置の実施例 3 は、本発明の実施の形態に係るテキスト情報作成装置が出力する重要節を強調表示したテキストまたは本発明の実施の形態に係るテキスト情報作成装置が出力する要約文に現れる全単語を検索キーまたは検索用のクエリとして使用する検索装置である。

【0182】

この検索の仕方については、特に言及しないが、たとえば、実施例 1 に係る事例寄せ装置を使用して、キーとなる検索テキストに関して事例寄せを行い、この事例寄せによってまとめられたテキストの集合の中から、キーとなる検索テキストの内容との距離が近い順にユーザが決めた数までのテキストを表示する方法などが考えられる。

【0183】

実施例3に係る検索装置の具体例としては、たとえば、図23のテキスト1の内容がキーとなる検索テキストだった場合、テキスト1の要約文と類似している、もしくはこの要約文に含まれる調理実習、鴨鍋、野菜グラタン、料理、作り方、教えてなどの単語を多く含む要約文を得ることができるテキスト3の質問事例を検索可能とする検索装置が考えられる。

【0184】

このような実施例3に係る検索装置は、質問事例と質問事例への回答とが対応して記載されているようなDBから、質問事例への回答を抽出したい場合などに有効である。

【0185】

以上説明したように、本発明の実施の形態に係るテキスト情報作成装置によれば、テキストの内容に関連する節をテキストから抽出できるため、検索や事例寄せを行う際にテキストの内容を容易に理解でき、検索や事例寄せの精度が高くなる。

【0186】

また、本発明の実施の形態に係るテキスト情報作成装置によれば、コーパスが使用されるため、単純に談話構造解析の結果のみを使用したのでは内容の類似性を強調できないテキストについても、検索や事例寄せの精度が向上する。すなわち、本発明の実施の形態に係るテキスト情報作成装置は、コーパス内に内容の類似性を強調できないテキストがあれば1テキスト以上見つけだし、見つけ出した1テキスト以上のテキストに含まれる単語の文字面などの談話構造解析結果以外の属性も使用するため、談話構造解析ではうまくいかなかったテキストを用いて検索や事例寄せを行う場合にも、検索や事例寄せの精度を向上させる。

【0187】

また、上述したように、従来技術4では、テンプレート作成をするためにはコーパスもしくは同種の表を作成後、作成されたコーパスまたは同種の表に含まれるテキスト自身の形式的な特徴や重要度の高い節の形式的な特徴を人手で捉えてテンプレートの形式やテキストもしくは節からテンプレートへの変換ルールを作

成しなくなければならないが、本発明の実施の形態に係るテキスト情報作成装置によれば、コーパスと談話構造解析ルールとを作成するだけで済む。

【0188】

したがって、本発明の実施の形態に係るテキスト情報作成装置によれば、談話構造解析ルールを作成するコストを考えた場合、テンプレートを作成する手法と比較して必要なコストは増えていない。

【0189】

また、本発明の実施の形態に係るテキスト情報作成装置によれば、談話構造解析ルールは文末表現が似ているテキストならどのような分野のテキストにも適応できるためいくつかの分野にて使用することができ、総合的にはテンプレートを作成する手法よりもコストの減少が可能である。

【0190】

さらに、本発明の実施の形態に係るテキスト情報作成装置によれば、コーパスの量が少ない場合や談話構造解析が失敗した場合にも要約の実行が可能であり、この点でも、テンプレート作成を行う手法よりも優れている。

【0191】

【発明の効果】

以上説明したように、本発明によれば、テキストの内容に強く関連する節を過剰に膨大な量の人手によるコストを必要とせずにテキストから抽出し、この抽出した節を用いてテキストについての情報を作成することができる。

【0192】

したがって、本発明によれば、テキストの検索や事例寄せなどのテキストの類似性の検討を必要とする作業や処理などにおいて、内容面で類似するテキストを探し出すことができる情報を容易に作成することができる。

【図面の簡単な説明】

【図1】

本発明の実施の形態に係るテキスト情報作成装置の概要図である。

【図2】

本発明の実施の形態に係るテキスト情報作成装置における処理を説明するフロ

ーチャートである。

【図 3】

図 2 のステップ S 2 - 1 で行われる前処理を説明するフローチャートである。

【図 4】

図 3 のステップ S 3 - 1 で行われる初期属性集合を構成する属性の作成を説明するフローチャートである。

【図 5】

図 3 のステップ S 3 - 4 で行われる、単語属性作成部による単語属性の作成に関する処理を説明するフローチャートである。

【図 6】

図 3 のステップ 3 - 8 で行われる余分属性削除部の処理を説明するフローチャートである。

【図 7】

図 3 のステップ S 3 - 7 で行われる最終確認を説明するフローチャートである。

【図 8】

重要度推定部の処理を説明するフローチャートである。

【図 9】

重要節決定部の処理を説明するフローチャートである。

【図 10】

談話構造ルール DB の概念図である。

【図 11】

属性集合用 DB の概念図である。

【図 12】

コーパス DB の概念図である。

【図 13】

結果 DB の概念図である。

【図 14】

重要度用 DB の概念図である。

【図 15】

談話構造解析をする前のテキストを示す図である。

【図 16】

談話構造解析が行われた図 15 のテキストを示す図である。

【図 17】

本発明の実施の形態に係るテキスト情報作成装置から出力され得る要約文の一例を示す図である。

【図 18】

本発明の実施の形態に係るテキスト情報作成装置から出力され得る重要節が強調表示されたテキストの一例を示す図である。

【図 19】

単語属性を加えていない場合の属性集合である初期属性集合を最終属性集合として用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。

【図 20】

単語属性を加えた場合の属性集合を最終属性集合として用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。

【図 21】

余分含属性集合を最終属性集合として用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。

【図 22】

余分除属性集合を最終属性集合として用いた場合の、コーパス中のテキスト ID が 2 のテキストを重要度決定器に入力した際の各節と各重要度とを示す図である。

【図 23】

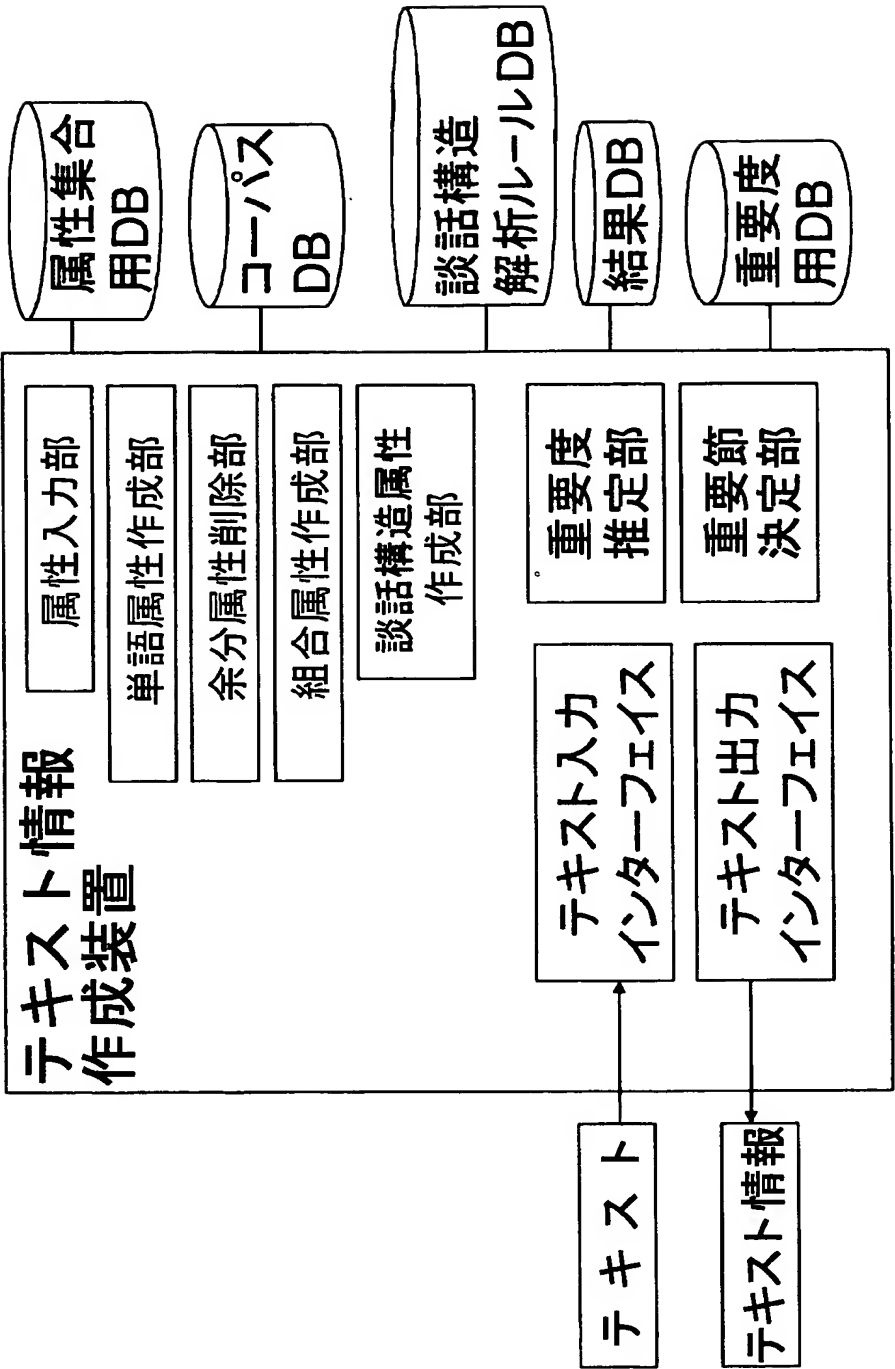
従来技術を説明するためのテキスト群を示す図である。

【書類名】

図面

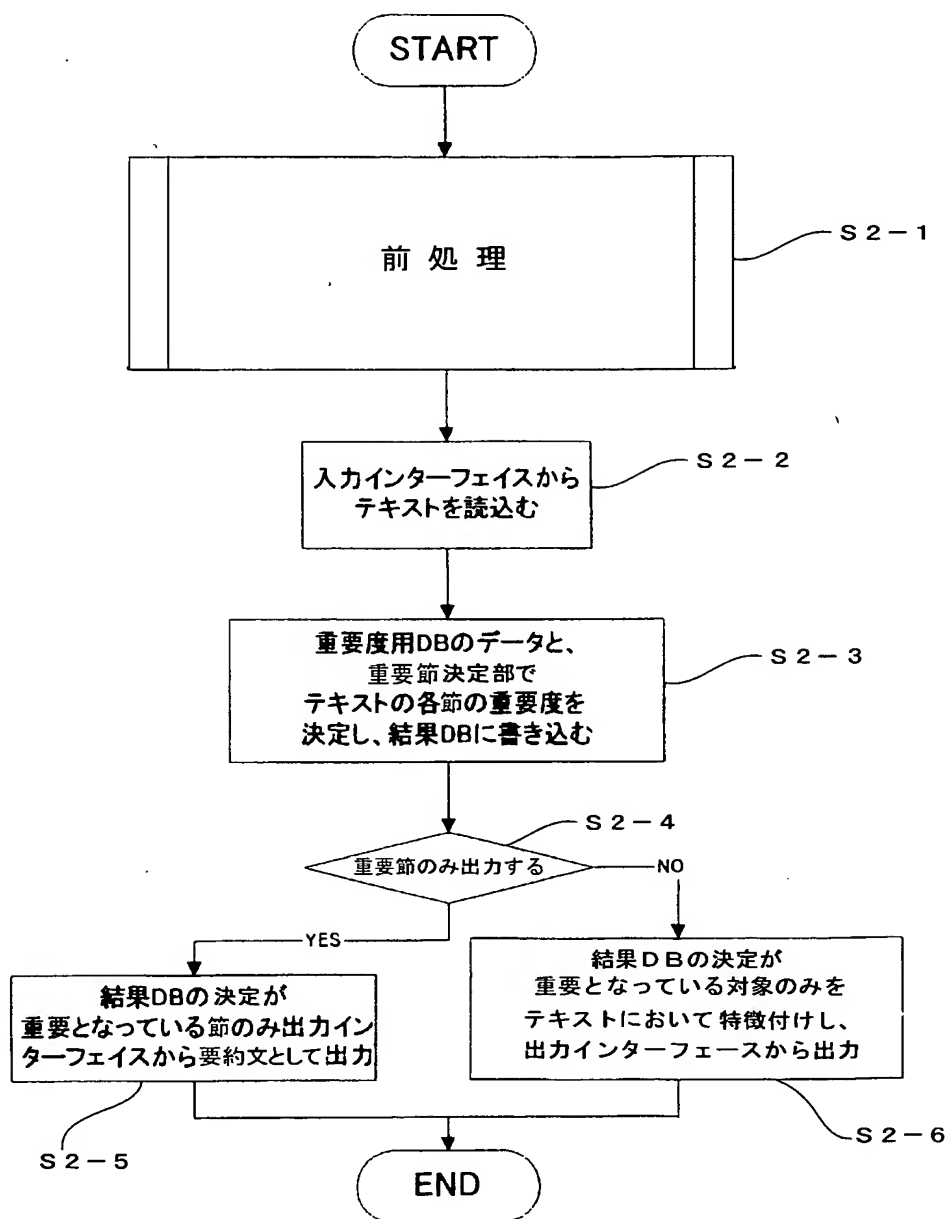
【図1】

図1 本発明の実施に係るテキスト情報作成装置の概要図



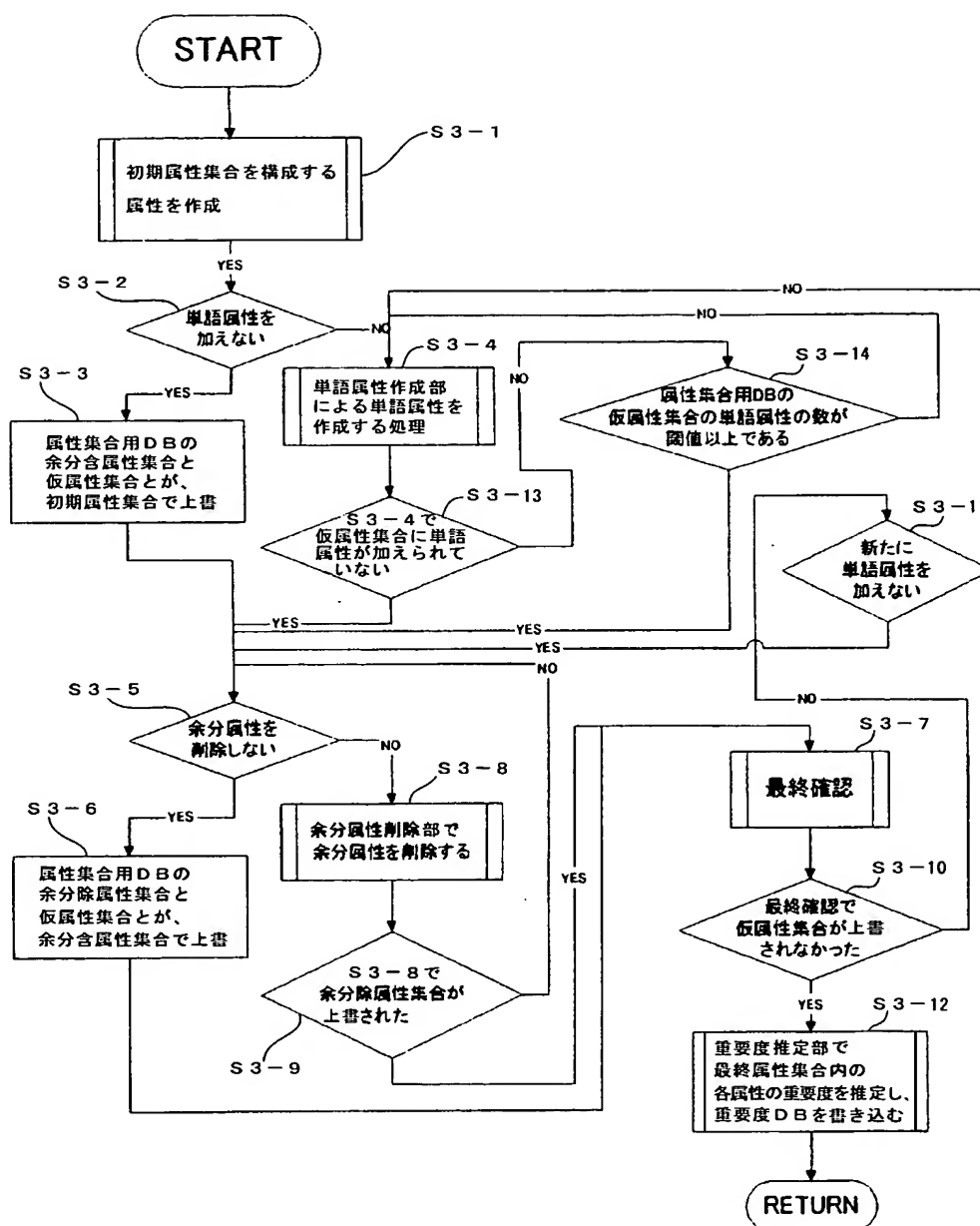
【図 2】

図 2 本発明の実施の形態に係るテキスト情報作成装置における処理を説明するフローチャート



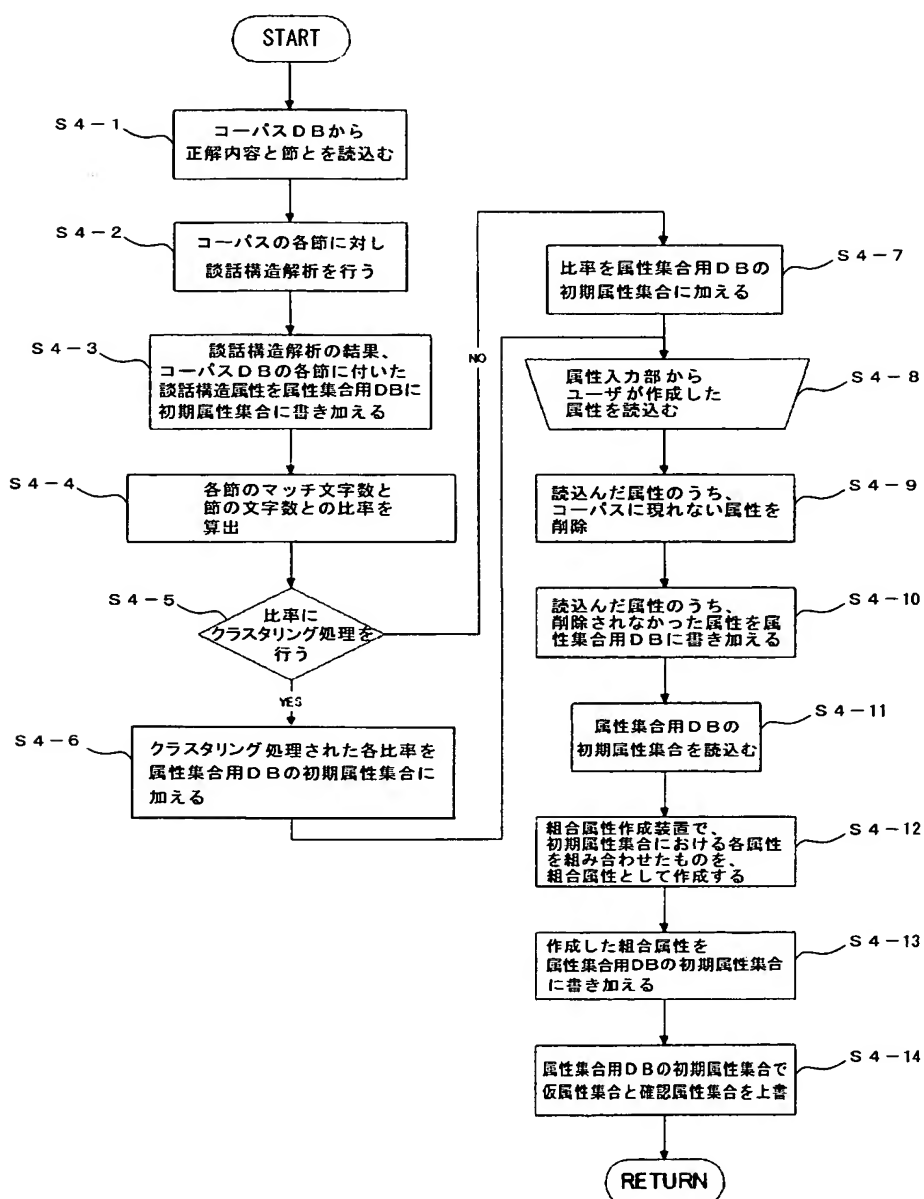
【図 3】

図 3 前処理を説明するフローチャート



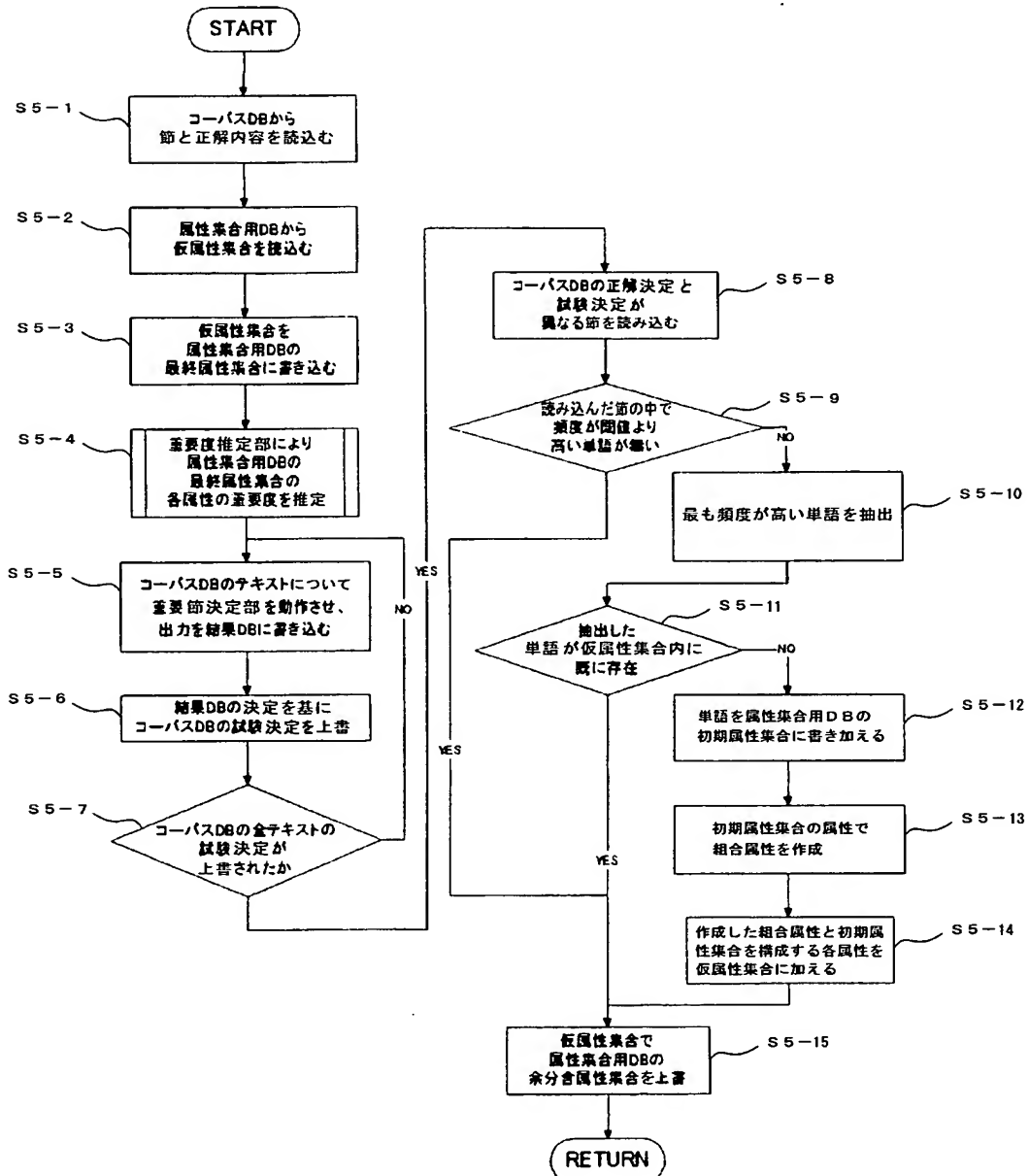
【図 4】

図 4 初期属性集合を構成する属性の作成を説明するフローチャート



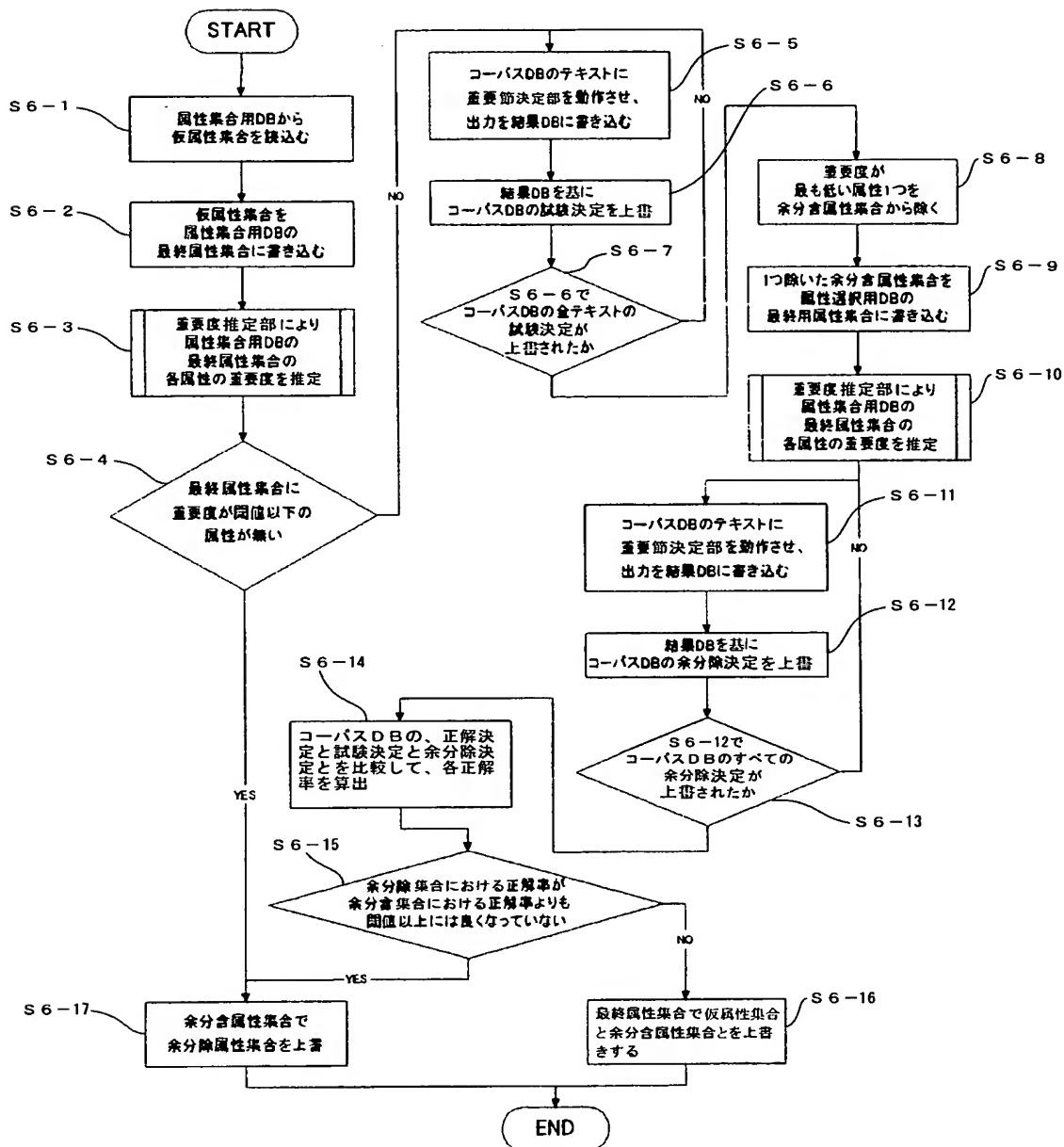
【図 5】

図5 単語属性作成部による単語属性の作成に関する処理を説明するフローチャート



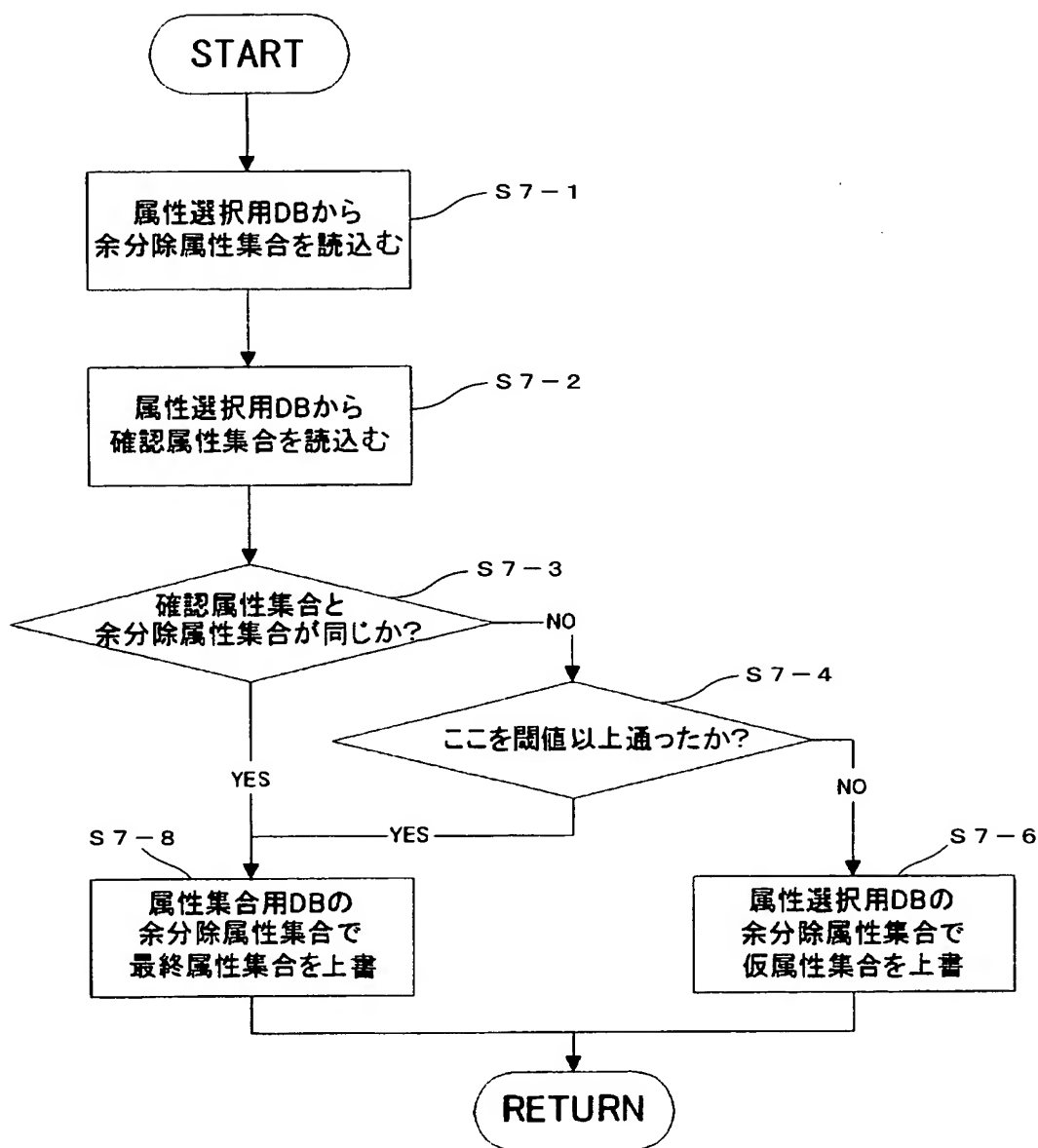
【図 6】

図 6 余分属性削除部による処理を説明するフローチャート

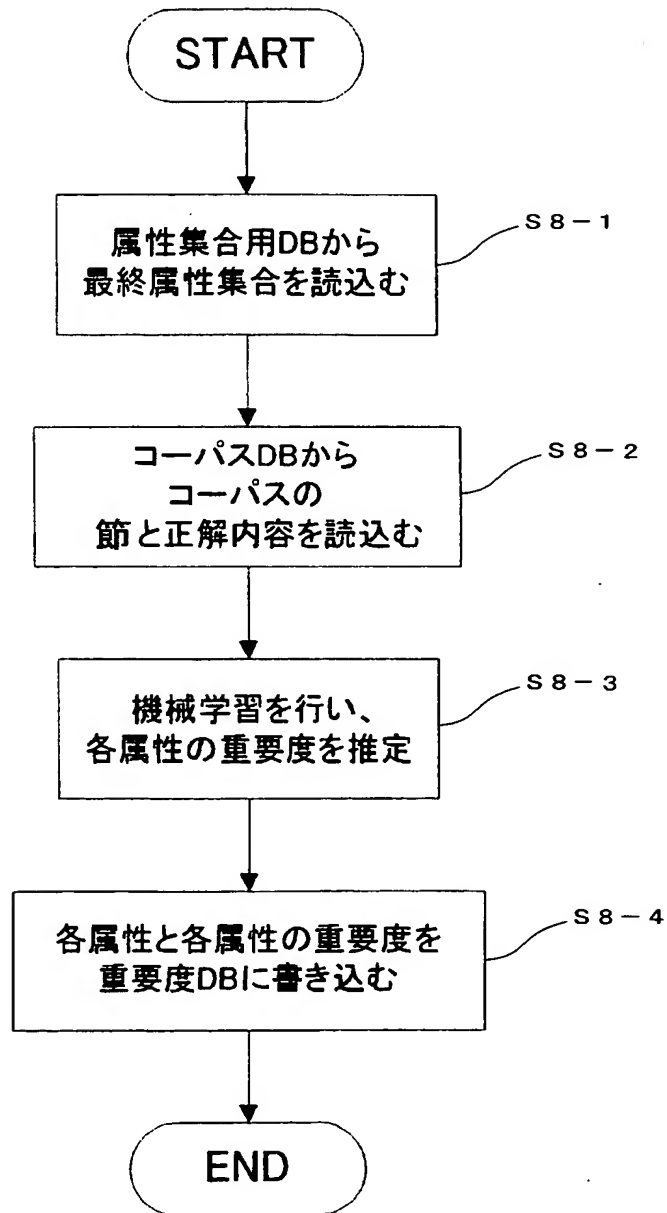


【図 7】

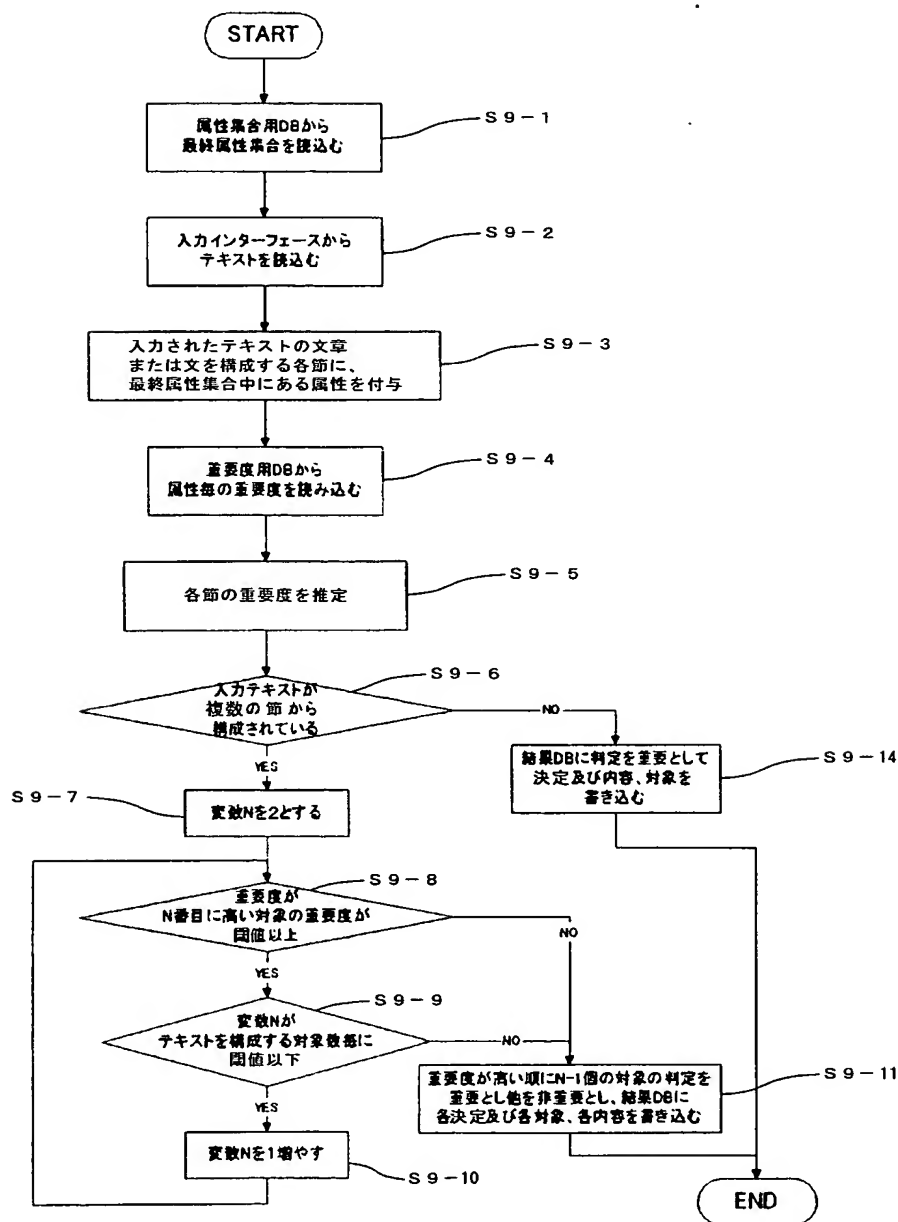
図 7 最終確認を説明するフローチャート



【図 8】

図 8 重要度推定部の処理を説明する
フローチャート

【図 9】

図 9 重要節決定部の処理を説明する
フローチャート

【図 10】

図10 談話構造ルールDB

談話構造	マッチングパターン
状態説明	しています
要求	教えて欲しい
希望	したいのです
要求	してくださいますか
質問	くださいますか
質問	可能ですか
要望	もらいたいです
礼	ありがとうございました
開始挨拶	はじめまして
開始挨拶	こんにちは
終了挨拶	よろしく願っています
.	.
.	.
.	.

【図 11】

図11属性集合用DB

属性の種類 属性集合	談話構造 属性	節長さ 比率属性	人為属性	組合属性	単語 属性
初期 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が2倍、比 率が3倍、比率が4 倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	
仮 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が2倍、比 率が3倍、比率が4 倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	PC、突然、Me、MB、 MHz、止り、落、設定、 使用、メール、……
確認 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が2倍、比 率が3倍、比率が4 倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	PC、突然、Me、MB、 MHz、止り、落、設定、 使用、メール、……
余分含 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が3倍、比 率が4倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	PC、突然、Me、MB、 MHz、止り、落、設定、 使用、メール、……
余分除 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が3倍、比 率が4倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	PC、突然、Me、MHz、 止り、落、設定、使用、 メール、……
最終 属性集合	質問、現状説明、 行動説明、礼、 行動結果説明、 要望……	構造解析がされてい ない、比率が3倍、比 率が4倍……	解決という文字がある、 位置が全体の上側、 位置が全体の下 側……	比率が2倍以上且解 決の文字、談話構造 が質問で2倍以 下、……	PC、突然、Me、MHz、 止り、落、設定、使用、 メール、……

【図 1 2】

図12コーパスDB

テキストID	節ID	正解内容	正解決定	試験決定	余分除決定	節
1	1	節ごとの 重要度 =1.0	重要	重要	重要	NETBUIプロトコルを追加したい
1	2	節ごとの 重要度 =0.0	非重要	非重要	非重要	追加するにはどうしたら良いか教えて欲しい
1	3	節ごとの 重要度 =1.0	重要	重要	重要	追加に必要なものと追加の仕方を送ってほしい
2	1	節ごとの 重要度 =0.0	非重要	非重要	非重要	はじめてメールいたします。
2	2	節ごとの 重要度 =0.0	非重要	非重要	非重要	そちらで販売のPCを使用しております。
2	3	節ごとの 重要度 =1.0	非重要	非重要	非重要	確かにPCを消してから寝ておりました。
2	4	節ごとの 重要度 =1.0	重要	重要	重要	夜中突然パソコンが立ちあがってしまいました。
.
.
.

【図 13】

図13結果DB

節ID	内容	決定	節
1	重要度 =0.01	非重要	はじめてメールいたします。
2	重要度 =0.24	非重要	そちらで販売のPCを使用しております。
3	重要度 =0.31	非重要	確かにPCを消してから寝ておりました。
4	重要度 =0.89	重要	ところが夜中突然パソコンが立ちあがってしまいました。
5	重要度 =0.21	非重要	以前はこんな事は有りませんでした。
6	重要度 =0.93	重要	どう設定したらこんな事が起きなくなるようになるかおしえて下さい。
7	重要度 =0.00	非重要	よろしく願います。

【図 14】

図14重要度用DB

属性の種類	属性	属性の重要度
談話構造属性	談話構造が質問	0.922
談話構造属性	談話構造が現状説明	0.521
談話構造属性	談話構造が行動説明	0.422
	.	.
	.	.
対象長さ比率属性	比率が2倍	0.501
対象長さ比率属性	比率が3倍	0.671
	.	.
	.	.
人為属性	現れる位置が全体の上側	0.925
	.	.
	.	.
組合属性	比率が2倍で談話構造が質問	0.525
	.	.
	.	.
単語属性	単語「PC」	0.337
	.	.
	.	.

【図 15】

図15 入力されたテキスト

はじめまして。

そちらのPCをかってネットワークを組もうとしています。
NETBUIプロトコルが無いので追加したいのですが……
追加するにはどうしたら良いか教えて欲しい。
追加に必要なものがありましたら、送ってもらいたいです。
よろしくお願いします。

【図 16】

図16談話構造解析の結果

<開始挨拶 マッチ数=5文字>はじめて</開始挨拶>。
<現状説明 マッチ数=5文字>そちらのPCをかってネットワーク
を組もうとしています</現状説明>。
<希望 マッチ数=6文字> NETBUIプロトコルが無いので追加
したいのですが…… </希望>
<要求 マッチ数=6文字>追加するにはどうしたら良いか教え
て欲しい</要求>。
<要求 マッチ数=5文字>追加に必要なものがありましたら、
送ってもらいたいです</要求>。
<終了挨拶 マッチ数=8文字>よろしくお願いいたします </終
了挨拶>。

【図 1 7】

図17 要約文

NETBUIプロトコルが無いので追加したいのですが……
追加に必要なものがありましたら、送ってもらいたいです。

【図 18】

図18 強調表示

はじめまして。
そちらのPCをかってネットワークを組もうとしています。
NETBUIプロトコルが無いので追加したいのですが……
追加するにはどうしたら良いか教えて欲しい。
追加に必要なものがありましたら、送ってもらいたいです。
よろしくお願いします。

【図 1 9】

図19 初期属性集合に属する属性を使用した
場合の各節の重要度

節ID	内容	節
1	重要度=0.43	はじめてメールいたします。
2	重要度=0.30	そちらで販売のPCを使用しております。
3	重要度=0.29	確かにPCを消してから寝ておりました。
4	重要度=0.35	ところが夜中突然パソコンが立ちあがってしまいました。
5	重要度=0.40	以前はこんな事は有りませんでした。
6	重要度=0.50	どう設定したらこんな事が起きなくなるようになるかおしえて下さい。
7	重要度=0.00	よろしく願います。

【図 2 0】

図20 単語属性を加えた場合の属性集合に属する属性を使用した場合の各節の重要度

節ID	内容	節
1	重要度=0.23	はじめてメールいたします。
2	重要度=0.63	そちらで販売のPCを使用しております。
3	重要度=0.50	確かにPCを消してから寝ておりました。
4	重要度=0.61	ところが夜中突然パソコンが立ちあがってしまいました。
5	重要度=0.20	以前はこんな事は有りませんでした。
6	重要度=0.64	どう設定したらこんな事が起きなくなるようになるかおしえて下さい。
7	重要度=0.00	よろしく願います。

【図 2 1】

図21 余分含属性集合に属する属性を使用した
場合の各節の重要度

節ID	内容	節
1	重要度=0.01	はじめてメールいたします。
2	重要度=0.23	そちらで販売のPCを使用しております。
3	重要度=0.31	確かにPCを消してから寝ておりました。
4	重要度=0.89	ところが夜中突然パソコンが立ちあがってしまいました。
5	重要度=0.22	以前はこんな事は有りませんでした。
6	重要度=0.94	どう設定したらこんな事が起きなくなるようになるかおしえて下さい。
7	重要度=0.00	よろしく願います。

【図 22】

図22 余分除集合に属する属性を使用した 場合の各節の重要度

節ID	内容	節
1	重要度=0.01	はじめてメールいたします。
2	重要度=0.24	そちらで販売のPCを使用しております。
3	重要度=0.31	確かにPCを消してから寝ておりました。
4	重要度=0.89	ところが夜中突然パソコンが立ちあがってしまいました。
5	重要度=0.21	以前はこんな事は有りませんでした。
6	重要度=0.93	どう設定したらこんな事が起きなくなるようになるかおしえて下さい。
7	重要度=0.00	よろしく願います。

【図 23】

図23 従来技術における課題を説明するためのテキスト群

テキストID	テキストの内容
テキスト1	おはようございます。研修中で忙しい所すみません。調理実習の課題は鴨の肉を使った鍋か大根を使用したサラダでした。でも私は今回出された課題の料理を作る事ができません。料理実習の時間までに作り方を教えていただけたらと助かります。よろしく願いいたします。
テキスト2	おはようございます。プログラミング研修の課題は鴨のグラフィックを動かす事だったと思います。しかし私は今回参加できませんでした。次回までに上記のプログラムを作っておく必要があるかを教えていただけると助かります。よろしく願いいたします。
テキスト3	大根サラダと鴨鍋の作り方が知りたい。ついでに秘訣を伝授いただきたければうれしいのですが……だめですか？

【書類名】 要約書

【要約】

【目的】 テキストの内容に強く関連する単語等を過剰に膨大な量の手によるコストを必要とせずにテキストから抽出し、この抽出した単語等を用いてテキストについての情報を作成することを目的とする。

【構成】 テキスト情報作成装置であって、人為属性を入力される属性入力部と、談話構造属性および節長さ比率属性を作成する談話構造属性作成部と、人為属性と談話構造属性および節長さ比率属性とを任意に組み合わせた組合属性を作成する組合属性作成部と、各属性についてテキストの内容との相関を高める度合いを示す重要度をそれぞれ推定する重要度推定部と、テキスト入力インタフェースと、入力されたテキストにおける 1 以上の節の中から重要な節を各属性の重要度に基づいて決定する重要節決定部と、重要節決定部の決定に基づいて作成される入力されたテキストについての情報を出力するテキスト出力インタフェースと、を備えるように構成する。

【選択図】 図 1

特願 2003-069658

出願人履歴情報

識別番号

[000005223]

1. 変更年月日

1996年 3月26日

[変更理由]

住所変更

住 所

神奈川県川崎市中原区上小田中4丁目1番1号

氏 名

富士通株式会社

{2}